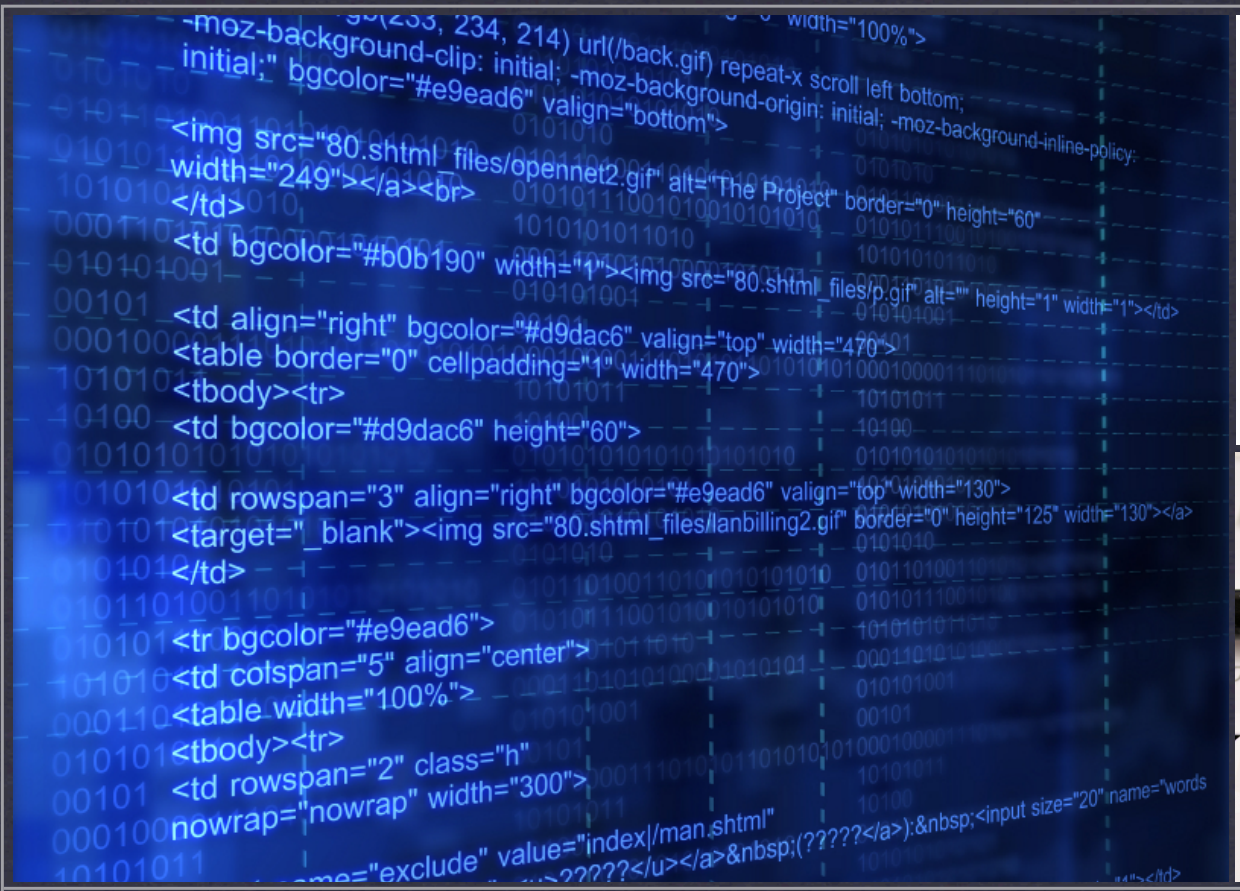


EDRN PUBLIC PORTAL

AND EDRN KNOWLEDGE ENVIRONMENT

Presented by Sean Kelly
EDRN Informatics Center
Jet Propulsion Laboratory

Division of
CANCER PREVENTION

JPL

Copyright © 2008 California Institute of Technology.
ALL RIGHTS RESERVED. U.S. Government sponsorship acknowledged.

Portal: what is it good for?

- Public *and private* face of EDRN
- First hit from Google
- Shows how taxpayer dollars are spent
- Encourages more sites to join in the revolution

Content management system

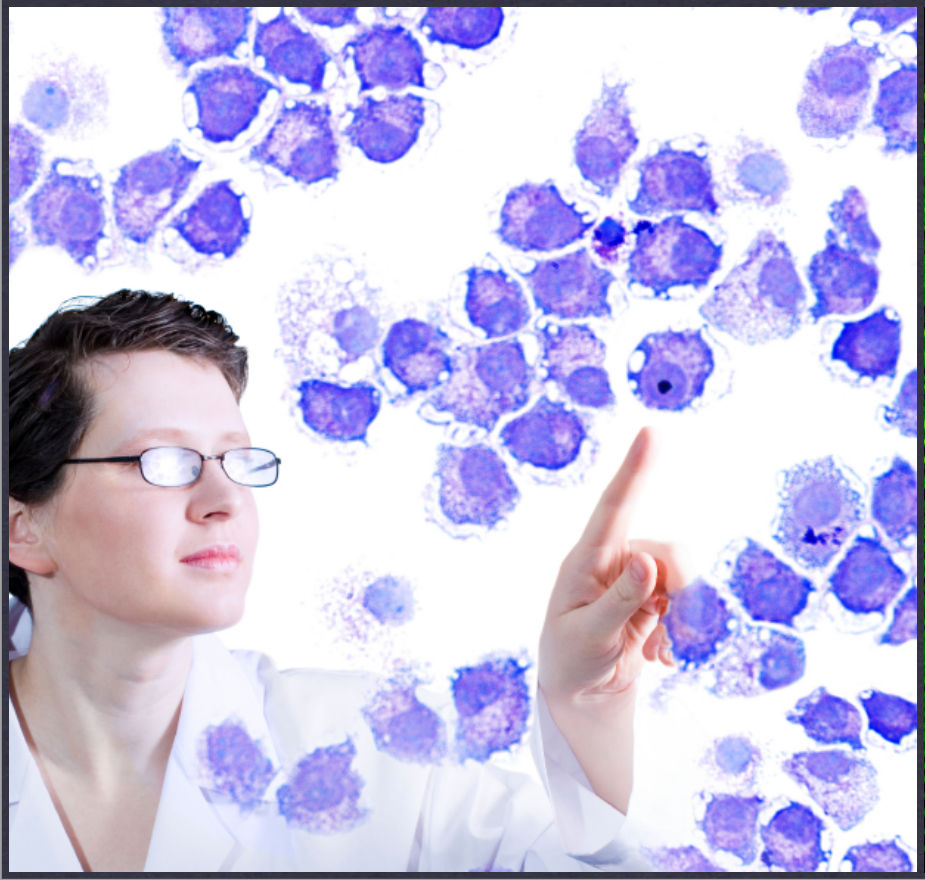
- Plone
  - #1 open-source CMS
  - Agile application server
  - Agile programming language
- Shockingly secure
- Search-engine friendly

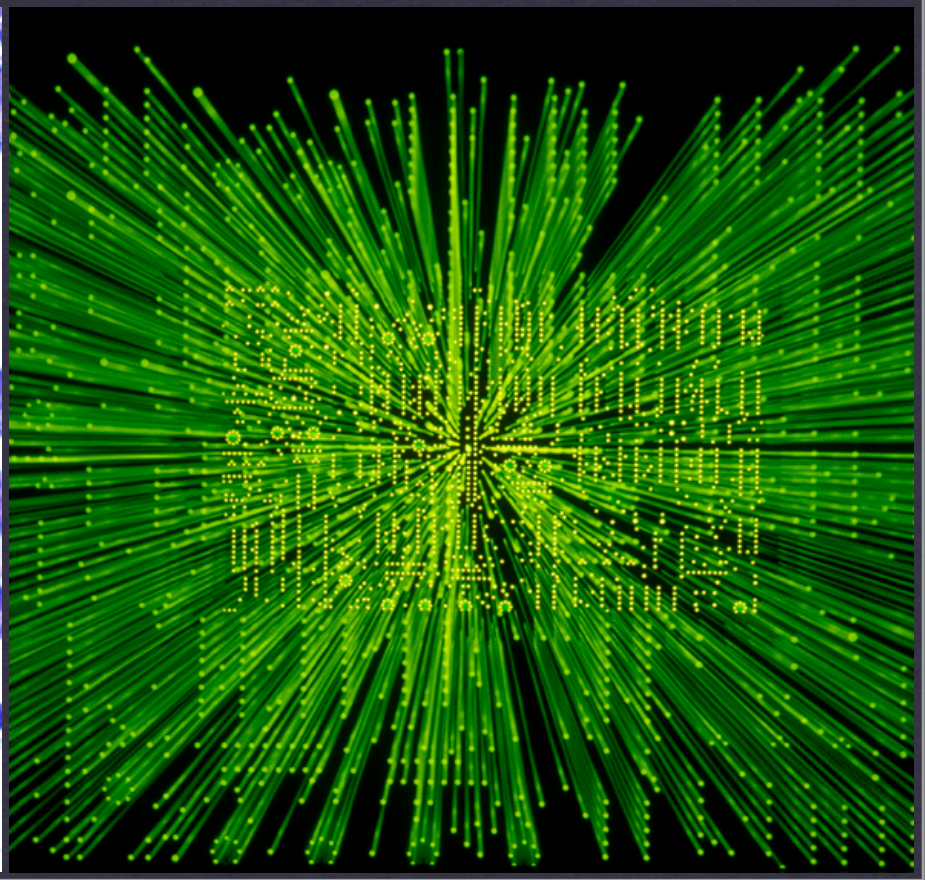
Flexibility

- Contributors
  - Who can create content
- Editors
  - Who can review and approve content
- Through the web
  - No HTML knowledge required

Domain specific

- Content tailored to EDRN's needs





THE ONTOLOGY

WHAT'S THE RELATIONSHIP?

Role of a CMS

- A content management system's job is to … manage content
- Various types of content
  - Web pages
  - Images
  - News items and events

Role of a CMS

- Workflow
  - State changes of content
  - Created → submitted for review → Reviewed by editor → Approved for publication → Published
    - → Retracted → Corrected → Re-approved → …

Role of a CMS

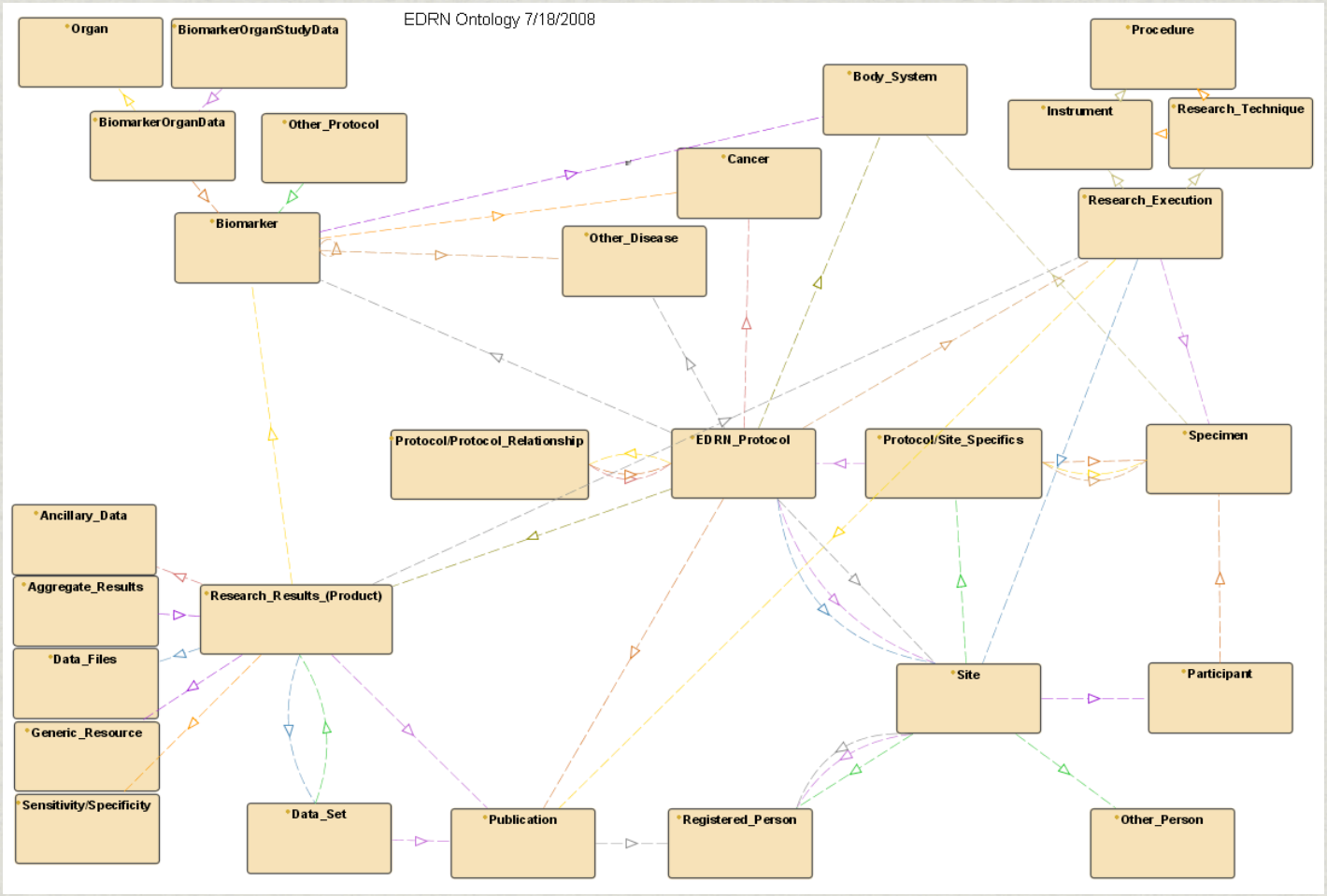
- Expand the definition of content
  - Studies, biomarkers, sites, specimens, publications, investigators
- Expand the definition of workflow
  - Collected → centrifuged → separated → alliquoted → frozen → …
  - Proposed → IRB-approved → funded → …
  - Phase 1 → 2 → 3 → 4 → 5

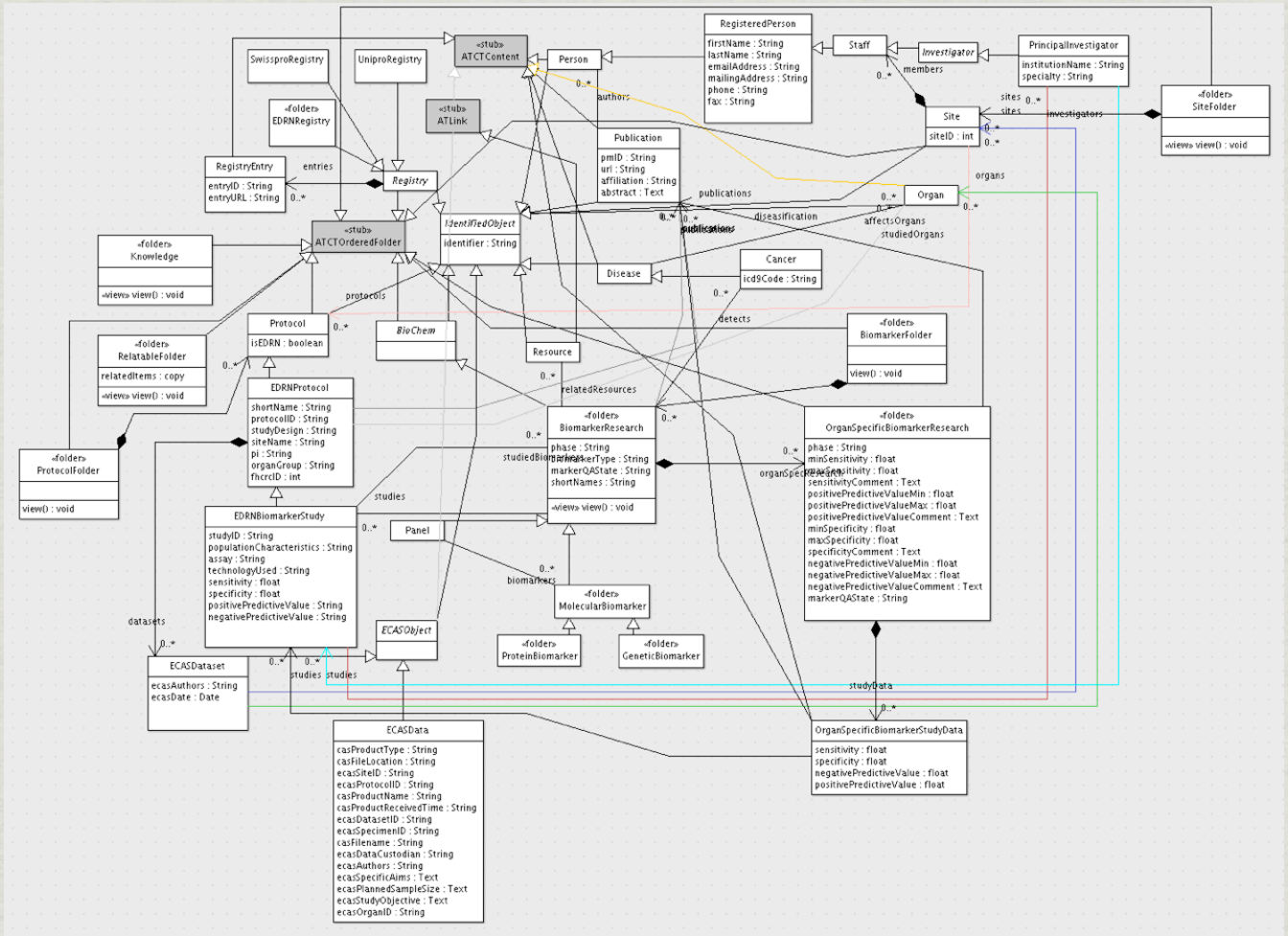
Role of a CMS

- New web application
  - Integrated with the portal
  - Searchable
  - Interlinked
  - Secure and robust

Portal + Ontology = Joy

- Direct mapping
  - For the most part
- A class in the ontology = a content type in the portal


EDRN Ontology 7/18/2008
Organ
BiomarkerOrganStudyData
BiomarkerOrganData
Other_Protocol
Biomarker
Body_System
Cancer
Other_Disease
Procedure
Instrument
Research_Technique
Research_Execution
Protocol/Protocol_Relationship
EDRN_Protocol
Protocol/Site_Specifics
Specimen
Ancillary_Data
Aggregate_Results
Data_Files
Generic_Resource
Sensitivity/Specificity
Research_Results_(Product)
Data_Set
Publication
Registered_Person
Site
Participant
Other_Person


«stub»
ATCTContent
Person
RegisteredPerson
firstName : String
lastName : String
emailAddress : String
mailingAddress : String
phone : String
fax : String
Staff
Investigator
PrincipalInvestigator
institutionName : String
specialty : String
members
authors
SwissproRegistry
UniproRegistry
«folder»
EDRNRegistry
«stub»
ATLink
Site
siteID : int
sites
investigators
«folder»
SiteFolder
«view» view() : void
Publication
pmID : String
url : String
affiliation : String
abstract : Text
RegistryEntry
entryID : String
entryURL : String
entries
Registry
publications
Organ
organs
affectsOrgans
studiedOrgans
diseasification
«stub»
ATCTOrderedFolder
IdentifiedObject
identifier : String
«folder»
Knowledge
«view» view() : void
Disease
Cancer
icd9Code : String
detects
protocols
Protocol
isEDRN : boolean
BioChem
Resource
«folder»
BiomarkerFolder
view() : void
«folder»
RelatableFolder
relatedItems : copy
«view» view() : void
EDRNProtocol
shortName : String
protocolID : String
studyDesign : String
siteName : String
pi : String
organGroup : String
fhcrcID : int
relatedResources
«folder»
ProtocolFolder
view() : void
«folder»
BiomarkerResearch
phase : String
markerType : String
markerQAState : String
shortNames : String
«view» view() : void
studiedBiomarkers
«folder»
OrganSpecificBiomarkerResearch
phase : String
minSensitivity : float
maxSensitivity : float
sensitivityComment : Text
positivePredictiveValueMin : float
positivePredictiveValueMax : float
positivePredictiveValueComment : Text
minSpecificity : float
maxSpecificity : float
specificityComment : Text
negativePredictiveValueMin : float
negativePredictiveValueMax : float
negativePredictiveValueComment : Text
markerQAState : String
organSpecificBiomarkers
studies
EDRNBiomarkerStudy
studyID : String
populationCharacteristics : String
assay : String
technologyUsed : String
sensitivity : float
specificity : float
positivePredictiveValue : String
negativePredictiveValue : String
Panel
biomarkers
«folder»
MolecularBiomarker
«folder»
ProteinBiomarker
«folder»
GeneticBiomarker
datasets
ECASObject
ECASDataset
ecasAuthors : String
ecasDate : Date
studies
studyData
ECASData
casProductType : String
casFileLocation : String
ecasSiteID : String
ecasProtocolID : String
casProductName : String
casProductReceivedTime : String
ecasDatasetID : String
ecasSpecimenID : String
casFilename : String
ecasDataCustodian : String
ecasAuthors : String
ecasSpecificAims : Text
ecasPlannedSampleSize : Text
ecasStudyObjective : Text
ecasOrganID : String
OrganSpecificBiomarkerStudyData
sensitivity : float
specificity : float
negativePredictiveValue : float
positivePredictiveValue : float

Synchronization

- Models are out of sync
  - One is ideal
  - Other is practical
    - Implementation-constrained

RDF

- A “language” for describing resources
  - Where a resource = anything
  - Where anything = has a URI
  - Where URI = Uniform Resource Identifier

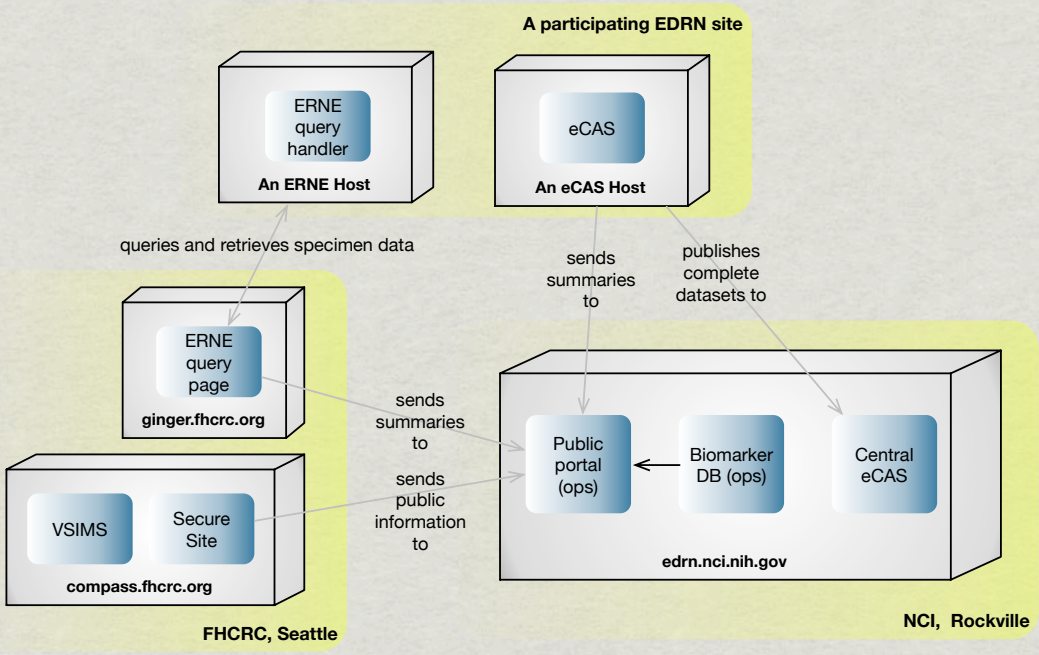
URI

- URLs are a kind of URI:
  - http://google.com/
- URNs are another kind of URI:
  - urn:isbn:817525766-0
  - urn:ssn:419-55-6113 (Dan Crichton's)

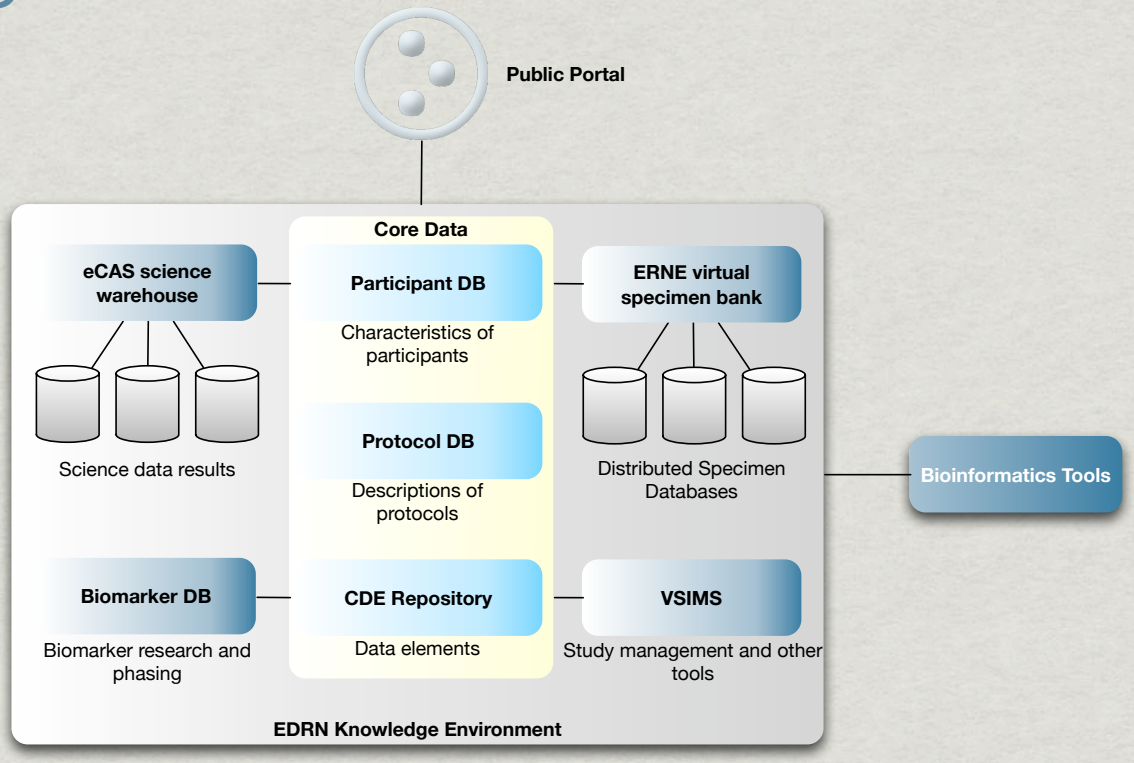
Send all EDRN knowledge

- To the public portal
- Via RDF
  - Biomarker database, VSIMS, ERNE
  - eCAS, eSIS, old scarab issue tracker, ASP pages
  - Any other SQL Server tables you find
  - Post-It™ notes, doodles on napkins …

Ingest data via RDF

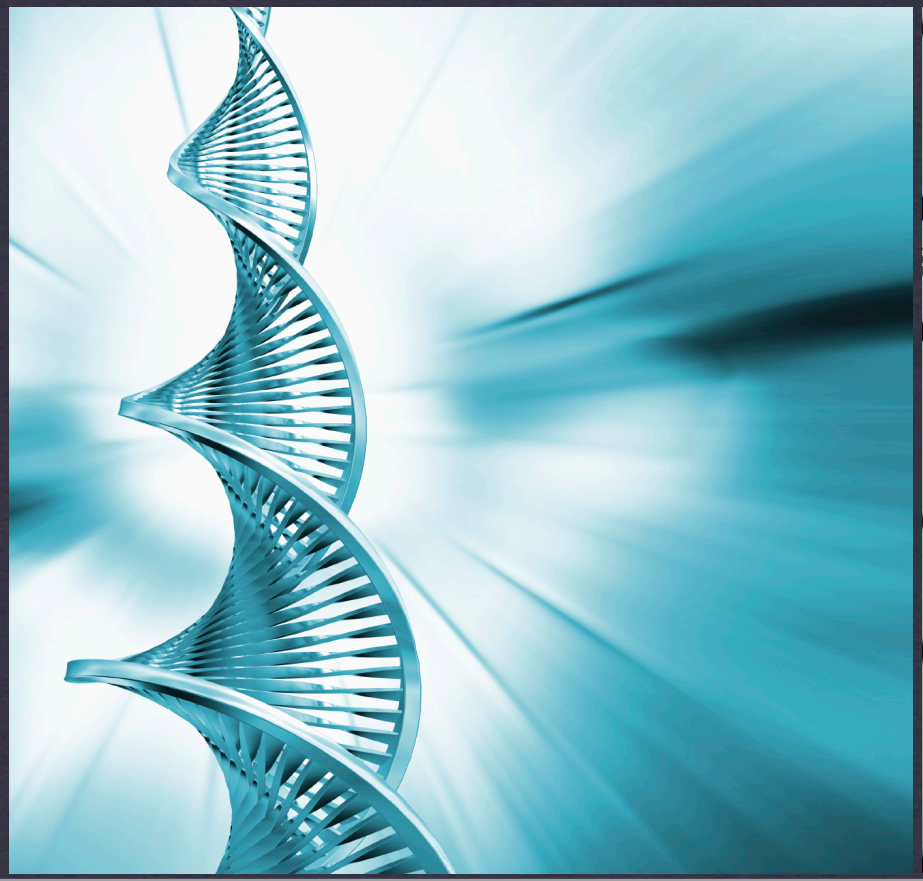


Realizes


Public Portal
Core Data
eCAS science warehouse
Participant DB
ERNE virtual specimen bank
Characteristics of participants
Science data results
Protocol DB
Descriptions of protocols
Distributed Specimen Databases
Bioinformatics Tools
Biomarker DB
CDE Repository
VSIMS
Biomarker research and phasing
Data elements
Study management and other tools
EDRN Knowledge Environment

DEMONSTRATION

JUST VISIT: HTTP://EDRN.JPL.NASA.GOV/



NEXT STEPS

WHERE DO WE GO FROM HERE?

Futures

- Deployment
  - Where will the portal finally live?
  - NCI? JPL?
    - Old Linux PC in my closet?

Futures

- Model synchronization
  - What working group must I join?
  - Can Protégé provide “views”?

Futures

- Plone 3 (We're on 2.5)
  - Exactly one year old yesterday
  - Released 2007-August-21
  - Just released: 3.2β2
- Major architectural changes
  - But a lot of new features and simplification

Futures

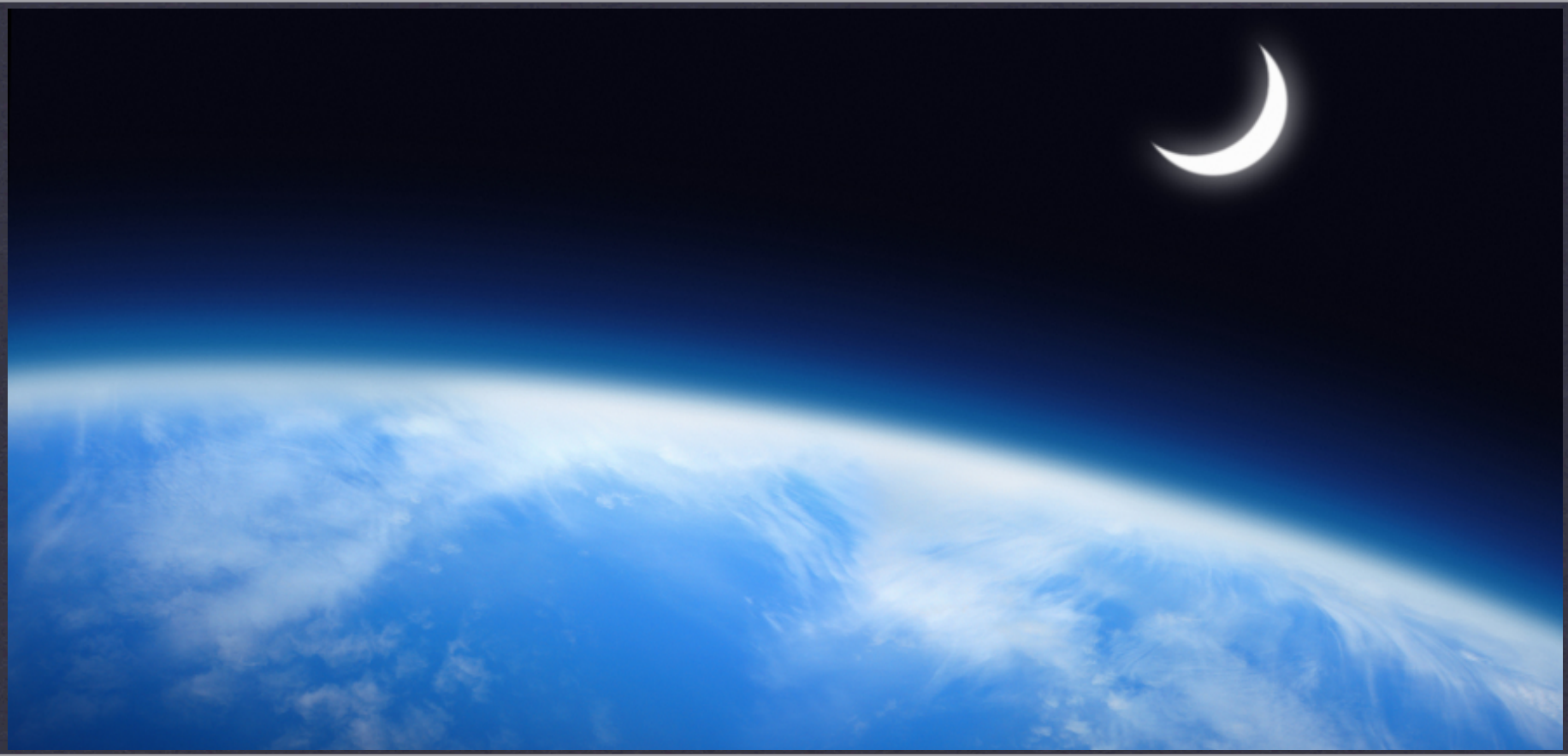
- More RDF
  - DMCC export all its knowledge
  - Implement routine updates
    - Cron-jobs

Futures

- Users and groups
  - Already authenticates against LDAP
  - Authorization
    - Define a policy for EDRN
      - Working groups, collaborative groups, etc.

Futures

- Anything else?



ANY QUESTIONS?

LET'S DISCUSS!